

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 July 2001 (12.07.2001)

PCT

(10) International Publication Number
WO 01/50672 A1

(51) International Patent Classification⁷: H04L 1/16, 1/18

(74) Agent: STUART, Michael, C.; Cohen, Pontani, Lieberman & Pavane, 551 Fifth Avenue, Suite 1210, New York, NY 10176 (US).

(21) International Application Number: PCT/IB00/01746

(22) International Filing Date:
24 November 2000 (24.11.2000)

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/477,457 4 January 2000 (04.01.2000) US

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant: NOKIA CORPORATION [FI/FI]; Keilalahdentie 4, FIN-02150 Espoo (FI).

(71) Applicant (*for LC only*): NOKIA INC. [US/US]; 6000 Connection Drive, Irving, TX 75039 (US).

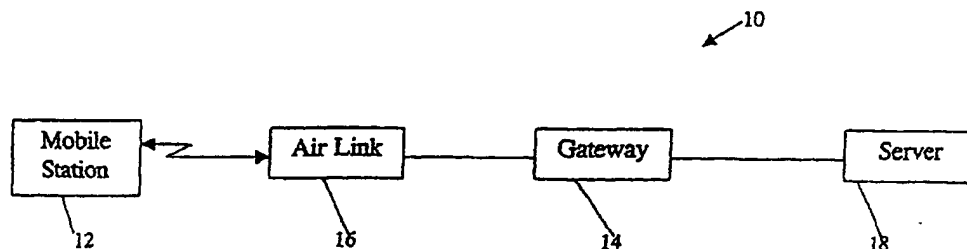
Published:

— With international search report.

(72) Inventors: GERENDAI, Magdolna; Lórántffy Zsuzsanna u.13, H-1043 Budapest (HU). TÓTH, Mihály; Kende Kanuth u. 67, H-1204 Budapest (HU). SZABÓ, Tamás; Vörösmarty u.69-71, V/2, H-1064 Budapest (HU).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DATA RETRANSMISSION METHOD



(57) Abstract: A method and system for large data transfer between a sender and a receiver. The sender transmits to the receiver a plurality of data packets in sequence. The time elapsed for each of the plurality of data packets after transmission of said each of the plurality of data packets is determined. The receiver transmits a message from the receiver to the sender notifying the sender that an identified one of the plurality of the data packets is missing. The sender retransmits to the receiver the identified one of the plurality of data packets only when the elapsed time determined for the identified one of the plurality of the data packets is greater than a predetermined time interval.

WO 01/50672 A1

DATA RETRANSMISSION METHOD

BACKGROUND OF THE INVENTION1. Field of the Invention

5 The present invention relates to data transmission in a wireless communication network and, more particularly, to a method and a system for transmitting data between a mobile station and a server in an efficient and reliable manner.

10

2. Description of the Related Art

 The Wireless Application Protocol (WAP) has been developed to address the needs of the mobile stations as these devices have severe physical, memory and processing constraints. WAP is applicable to a number of different systems including GSM-900, GSM-1800, GSM-1900, CDMA IS-95, TDMA IS-136, wide-band IS-95 and third generation systems such as IMT-2000, UMTS and W-CDMA.

20 Pursuant to WAP, there are three classes of transactions: Class 0 for unreliable one way requests, Class 1 for reliable one-way requests, and Class 2 for reliable two-way request-response transactions. A transaction is defined as a unit of interaction between an "initiator" (e.g., a sender or a receiver) and a "responder" (e.g., a sender or a receiver). A transaction begins with an invoke message generated by the sender. In class 2 transactions, the receiver replies with one result message that implicitly acknowledges the invoke message. If the receiver takes longer than a specified time interval to service the invoke message, the receiver may reply with a "hold on" acknowledgement message before transmitting the result message so as to avoid the retransmission of the invoke

30

message. The transaction ends when the sender receives the acknowledgement.

If the length of a message exceeds the maximum size specified by a bearer (e.g., a telephone company),
5 then before the message is sent, it is segmented into an ordered sequence of data packets. Each data packet has a packet sequence number (PSN) assigned thereto. Thus, for example, the first segmented data packet is assigned a packet sequence number of zero, the second data packet
10 is assigned a packet sequence number of one, the third data packet is assigned a packet sequence number of two, and so on. The data packets are transmitted individually or in groups, and then re-assembled by the receiver upon receipt. The maximum number of packets a
15 message can be segmented is about 256 packets, each packet having a maximum size of 1 to 2 Kbytes. Thus, the maximum size of a message is typically less than 0.5 Mbytes.

In the case where the data packets are
20 segmented into groups, the sender does not send any new packets belonging to the same transaction until receipt of the previous packet groups has been acknowledged. In other words, packet groups are sent according to a stop-and-wait protocol. Typically, the sender determines the
25 number of packets in each packet group and transmits the packets of a packet group in one batch. The last data packet of each packet group has a Group Trailer (GTR) flag set and is often referred to as the GTR packet. The last data packet of the last packet group of the entire
30 message has the Transmission Trailer (TTR) flag set and is also referred to as the TTR packet. The receiver stores all packets received and upon receipt of the GTR packet, the receiver checks whether it has received all packets belonging to that particular packet group. If

it has received all of the packets in that group, the receiver returns an acknowledgement message containing the packet sequence number of that GTR packet. If, however, the GTR or TTR packet is received and one or
5 more packets of the packet group are missing, the receiver waits for a period of time, e.g., $\frac{1}{2}$ the median round-trip time (RTT), before transmitting a negative acknowledgement (NACK) message containing the packet sequence numbers of the missing packets of that
10 particular packet group. If the sender does not receive an acknowledgement within a specified time interval after transmission of a packet group, it retransmits only the GTR or TTR packet of that packet group to the receiver.

15 Problems arise, however, as one or more of the acknowledgement or negative acknowledgement messages are lost during transmission due to, for example, handover errors or congestion in the wireless communication network, thereby resulting in the retransmission of data
20 packets that the receiver has already received.

SUMMARY OF THE INVENTION

An object of the present invention is to provide a more robust method and system of transmission
25 of data.

According to one aspect of the invention, a receiver transmits a negative acknowledgement message to a sender listing all previously missing data packets of a data message. The receiver implicitly acknowledges
30 received data packets by not listing them in the negative acknowledgement message.

According to another aspect of the invention, the sender retransmits a data packet only if the time

elapsed after the original transmission of that data packet is greater than a predetermined time period.

In one embodiment, the sender transmits to the receiver a plurality of data packets in sequence. The time elapsed for each of the plurality of data packets after transmission of said each of the plurality of data packets is determined. The receiver transmits a message from the receiver to the sender notifying the sender that an identified one of the plurality of the data packets is missing. The sender retransmits to the receiver the identified one of the plurality of data packets only when the elapsed time determined for the identified one of the plurality of the data packets is greater than a predetermined time interval.

Other objects and features of the present invention will become apparent from the following detailed description considered in conjunction with the accompanying drawings. It is to be understood, however, that the drawings are designed solely for purposes of illustration and not as a definition of the limits of the invention, for which reference should be made to the appended claims. It should be further understood that the drawings are not necessarily drawn to scale and that, unless otherwise indicated, they are merely intended to conceptually illustrate the structures and procedures described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings, wherein like reference characters denote similar elements:

Fig. 1 is a block diagram of a data transmission system configured in accordance with an embodiment of the present invention;

Fig. 2 is a segmented data message transmitted in the system of Fig. 1;

Fig. 3 is a message sequence diagram in accordance with one embodiment of the inventive method;
5 and

Fig. 4 is a message sequence diagram in accordance with another embodiment of the inventive method.

10 DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

Finnish Patent Application No. 999470, filed on November 17, 1999 and assigned to the assignee hereof, is incorporated herein by reference. That application discloses a new class of data transmission
15 for the WAP standard, in particular, for transmission of a data message of an arbitrary length.

Fig. 1 diagrammatically illustrates an exemplary data transmission system 10 comprising a mobile station 12 (e.g., a cellular phone) in wireless
20 communication with a gateway 14, through an air link 16 (that includes, for example, a base station and a switching control point). The gateway 14 is connected to a server 18 through a wide area network such as, for example, the Internet or Intranet. Data is transmitted
25 to and from the mobile station 12 in accordance with the inventive method described below.

Fig. 2 diagrammatically illustrates a segmented data message of a data stream. The data message comprises a plurality of data packet groups.
30 Each group comprises one or more data packets 20 with no flag sets and a data packet 22 with a GTR flag set. The GTR flag set indicates the last data packet of a data packet group. The last data packet 24 of the entire data message includes a TTR flag set. Each data packet

20 of a data message is associated with a unique identification number, e.g., a packet sequence number (PSN) so that a receiver (e.g., mobile station 12) can identify a missing data packet of the data message on the basis of the identification numbers. Preferably, the data packets are numbered consecutively (e.g., 0, 1, 2, 3 ...).

In one embodiment, a sender (e.g. gateway 14 or server 18) transmits to the receiver the data packets sequentially in accordance with the PSNs so that a data packet 20 with a lower PSN is transmitted prior to one with a higher PSN (e.g., a data packet with PSN=2 is transmitted before a data packet with PSN=3). If the receiver determines that a data packet 20 is missing (e.g., it has received data packet with PSN=3 but not PSN=1), it transmits to the sender a negative acknowledgement (NACK) message listing the missing packet (e.g., PSN=1). Preferably, the NACK message lists not only the missing packet(s) of the most current data packet group but also the missing packet(s) of all previous data packet groups. Thus, if a data packet with PSN=3 is missing from packet group 1 (and not yet received) and data packet with PSN=25 is missing from packet group 5, the NACK message will list both missing data packets (PSN=3 and PSN=25). This is particularly advantageous because even if the previous NACK message is lost, the sender will still be notified of the missing data packet(s) of previously sent packet groups. It is contemplated that the NACK message may be sent before or after the receiver receives a GTR data packet of each data packet group.

After receiving a NACK message from the receiver, the sender checks the time elapsed since the transmission of each missing data packet listed in the

NACK message. If the elapsed time for a missing data packet is less than a predetermined time interval (e.g., about one round trip time (RTT) for a data packet to travel between the sender and the receiver), the sender
5 does not retransmit the requested (missing) data packet. On the other hand, if the elapsed time for the missing data packet is greater than the predetermined time interval, then the sender retransmits the missing data packet.

10 Upon receipt of all data packets in a group, the receiver transmits an acknowledgement (ACK) message to the sender acknowledging receipt of all data packets for a group and then begins assembling the received data packets so as to reconstruct the data message. Similar
15 to the transmission of NACK messages, the receiver may transmit an ACK message after receipt of the GTR data packet of a group. The receiver may also transmit an ACK message acknowledging receipt of all data packets of a data message after receipt of the TTR data packet.
20 Alternatively, the receiver may transmit a NACK message listing no missing data packets for a particular group to thereby implicitly acknowledge receipt of all data packets in that particular group.

 In a particularly preferred embodiment, a NACK
25 message lists not only the missing data packets, but also the highest packet sequence number of the data packets received by the receiver. Using such a NACK message, the receiver can thus implicitly acknowledge receipt of all data packets with lower PSNs (except for
30 the listed missing packets). An example illustrating this NACK message is as follows. A sender transmits two groups of data packets: packets with $PSN = N, N+1, N+2(GTR)$; and packets with $PSN = N+3, N+4, N+5(GTR)$. The data packets with $PSN = N+2$ and $N+5$ include the GTR flag

sets and are thus the last data packets in their respective groups. The data packets with PSN=N and N+1 are lost during transmission. The receiver then transmits a NACK message (NACK_1) listing data packets
5 with PSN=N and N+1 as missing and data packet with PSN=N+2 as the last data packet (i.e., with the highest PSN) received by the receiver. NACK_1, however, is also lost during transmission. Meanwhile, the receiver continues to receive data packets of the next packet
10 group: data packets with PSN=N+3, N+4, and N+5. After determining data packets with PSN=N and N+1 are still missing, the receiver transmits a second NACK message (NACK_2) listing the last data packet received (e.g., data packet with PSN=N+5) and the still missing data
15 packets (data packets with PSN=N and N+1). The sender receives NACK_2 and recognizes that the receiver has received all data packets in the two groups of data packets, except data packets with PSN=N and N+1. The sender thereafter resends the missing data packets
20 (PSN=N and N+1). It is contemplated that this NACK message may be used in combination with an acknowledgement message to further enhance reliability of communication between the sender and the receiver.

Fig. 3 illustrates the steps of one embodiment
25 of the inventive method. A receiver transmits a data request to the air link 16, which in turn transmits the request to a sender. The sender receives the data request and transmits to the air link 16 data packets 20
30 with PSN=N, N+1, and N+2. The air link then transmits only data packets with PSN=N+1 and N+2 and loses the data packet 20 with PSN=N. The receiver determines that data packet with PSN=N is missing and transmits to the air link 16 a NACK message (NACK_1) listing the data packet with PSN=N. But the NACK message is also lost

and the air link 16 fails to forward the NACK message to the sender. The sender, still responding to the original data request, sends data packets 20 with PSN=N+3 and N+4 of another packet group. The air link 5 16 transmits to the receiver the data packets with PSN=N+3 and N+4. The receiver determines again that the data packet PSN=N is missing and sends another NACK message (NACK_2) listing the missing data packet to the air link 16, which successfully forwards the NACK_2 10 message to the sender. In response to NACK_2, the sender transmits the missing data packet with PSN=N to the receiver through the air link 16.

Fig. 4 illustrates another embodiment of the inventive method. The receiver transmits a request for 15 data to the air link 16, which in turn transmits the request to the sender. The sender then sends the data packets with PSN=N, N+1 to the air link 16. The air link loses data packet with PSN=N and transfers only data packet with PSN=N+1 to the receiver. The receiver 20 transmits to the air link 16 a negative acknowledgement message (NACK_1) listing data packet with PSN=N. The sender continues to send data packet with PSN=N+2 to the air link 16, which successfully transmits the data packet to the receiver. The receiver determines (before 25 or after receipt of a GTR packet) data packet with PSN=N is missing and thereafter transmits another negative acknowledge message (NACK_2). However, the air link finally transmits NACK_1 to the sender, which responds by sending the data packet with PSN=N to the air link 30 16. The air link 16 transfers the data packet with PSN=N to the receiver. After the sender transmits the data packet (PSN=N) within a time interval less than one RTT (i.e., the predetermined time interval), the sender receives NACK_2 requesting the same data packet (i.e.,

PSN=N). Since NACK_2 was received within a time interval less than one RTT, the sender ignores NACK_2 and refuses to send the requested data packet.

Thus, while there have shown and described and
5 pointed out fundamental novel features of the invention as applied to a preferred embodiment thereof, it will be understood that various omissions and substitutions and changes in the form and details of the devices illustrated, and in their operation, may be made by
10 those skilled in the art without departing from the spirit of the invention. For example, it is expressly intended that all combinations of those elements and/or method steps which perform substantially the same function in substantially the same way to achieve the
15 same results are within the scope of the invention. Moreover, it should be recognized that structures and/or elements and/or method steps shown and/or described in connection with any disclosed form or embodiment of the invention may be incorporated in any other disclosed or
20 described or suggested form or embodiment as a general matter of design choice. It is the intention, therefore, to be limited only as indicated by the scope of the claims appended hereto.

CLAIMS

What is claimed is:

1. A method of transmitting data between a
5 sender and a receiver, comprising in sequence the steps
of:

(a) transmitting a plurality of data packets
in sequence from the sender to the receiver in response
to the request;

10 (b) determining time elapsed for each of the
plurality of data packets after transmission of said
each of the plurality of data packets;

(c) determining by the receiver whether any
of the plurality of data packets is missing;

15 (d) transmitting a message from the receiver
to the sender notifying the sender that an identified
one of the plurality of the data packets is missing when
the receiver has determined at least one of the
plurality of data packets is missing; and

20 (e) retransmitting from the sender to the
receiver the identified one of the plurality of data
packets only when the elapsed time determined for the
identified one of the plurality of the data packets is
greater than a predetermined time interval.

25 2. The method of claim 1, wherein the
predetermined time interval is about one round trip time
for the identified one of the plurality of data packets
to travel between the sender and the receiver.

30 3. The method of claim 1, wherein each of
the plurality of data packets of a data message is
uniquely identified by a packet sequence number.

4. The method of claim 3, wherein the packet
sequence numbers are arranged in consecutive order.

5. The method of claim 1, wherein the data packets are arranged in groups, each group containing at least one data packet.

6. The method of claim 5, wherein the data
5 packets are transmitted in groups.

7. The method of claim 6, wherein each group includes a data packet with a GTR flag set for indicating the last data packet of a group.

8. The method of claim 7, wherein in step
10 (d) the sender is notified of all missing data packets in all previously sent groups.

9. The method of claim 1, wherein step (d) further comprises the step of notifying the sender of the last data packet received by the receiver.

10. The method of claim 7, wherein step (d)
15 further comprises the step of notifying the sender of the highest packet sequence number of the last data packet received by the receiver.

11. A system of transmitting data between a
20 sender and a receiver, comprising:

(a) means for transmitting a plurality of data packets in sequence from the sender to the receiver in response to the request;

(b) means for determining time elapsed for
25 each of the plurality of data packets after transmission of said each of the plurality of data packets;

(c) means for determining by the receiver whether any of the plurality of data packets is missing;

(d) means for transmitting a message from the
30 receiver to the sender notifying the sender that an identified one of the plurality of the data packets is missing when the receiver has determined at least one of the plurality of data packets is missing; and

(e) means for retransmitting from the sender to the receiver the identified one of the plurality of data packets only when the elapsed time determined for the identified one of the plurality of the data packets
5 is greater than a predetermined time interval.

12. The system of claim 11, wherein the sender one of gateway and a server.

13. The system of claim 11, wherein the receiver is a mobile station.

10 14. The system of claim 13, wherein said mobile station is a mobile phone.

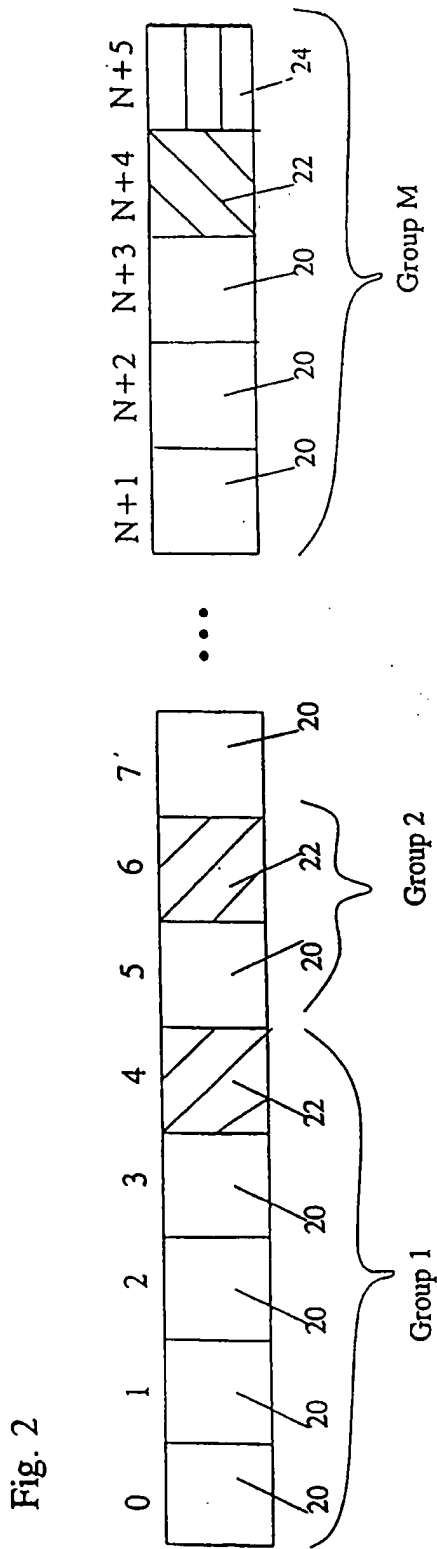
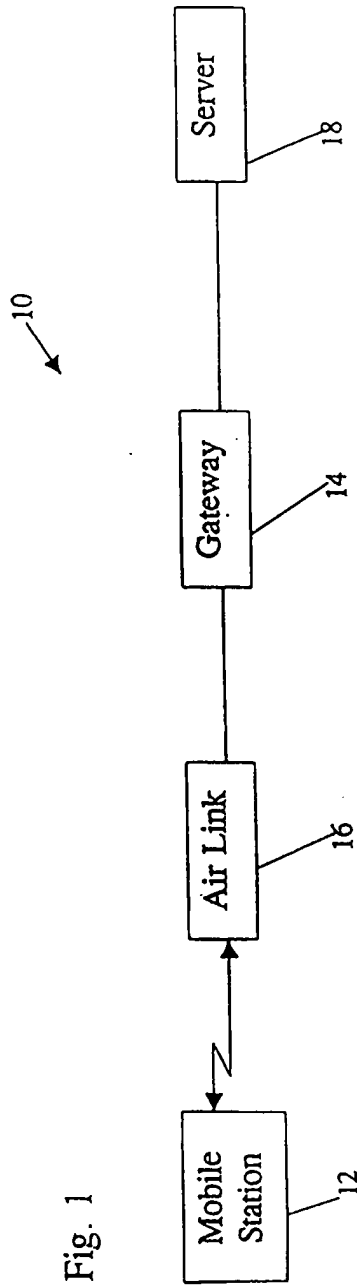


Fig. 3

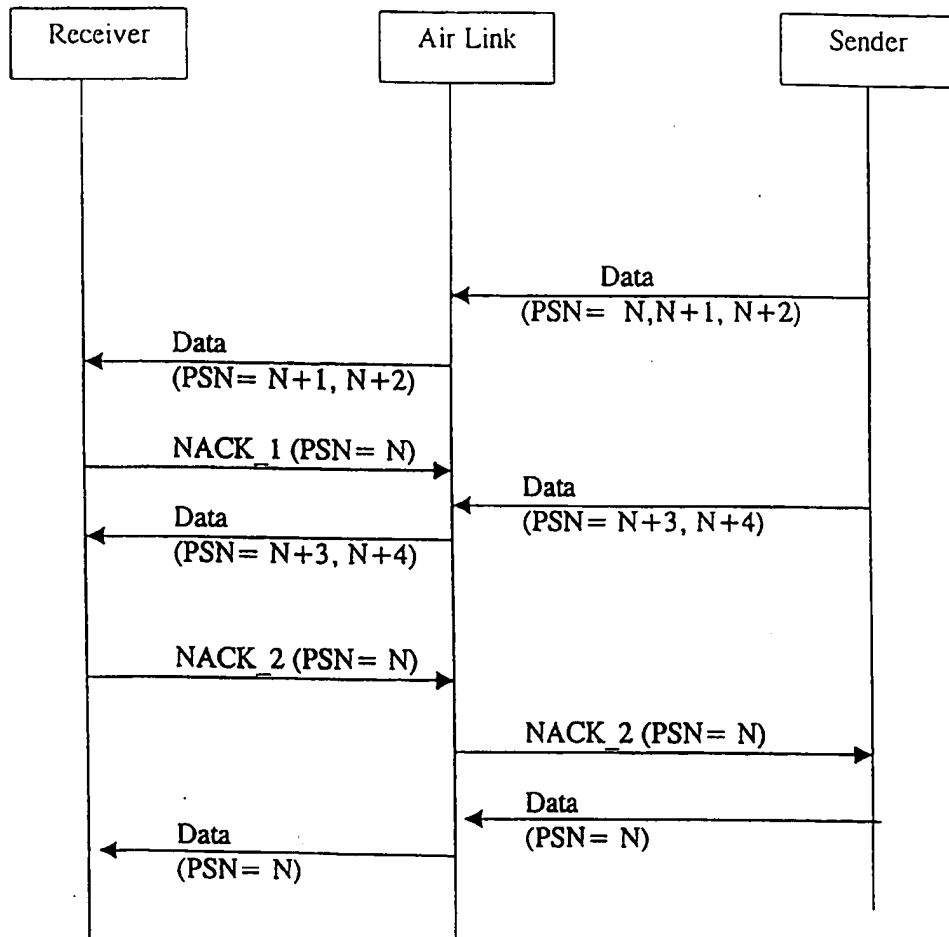
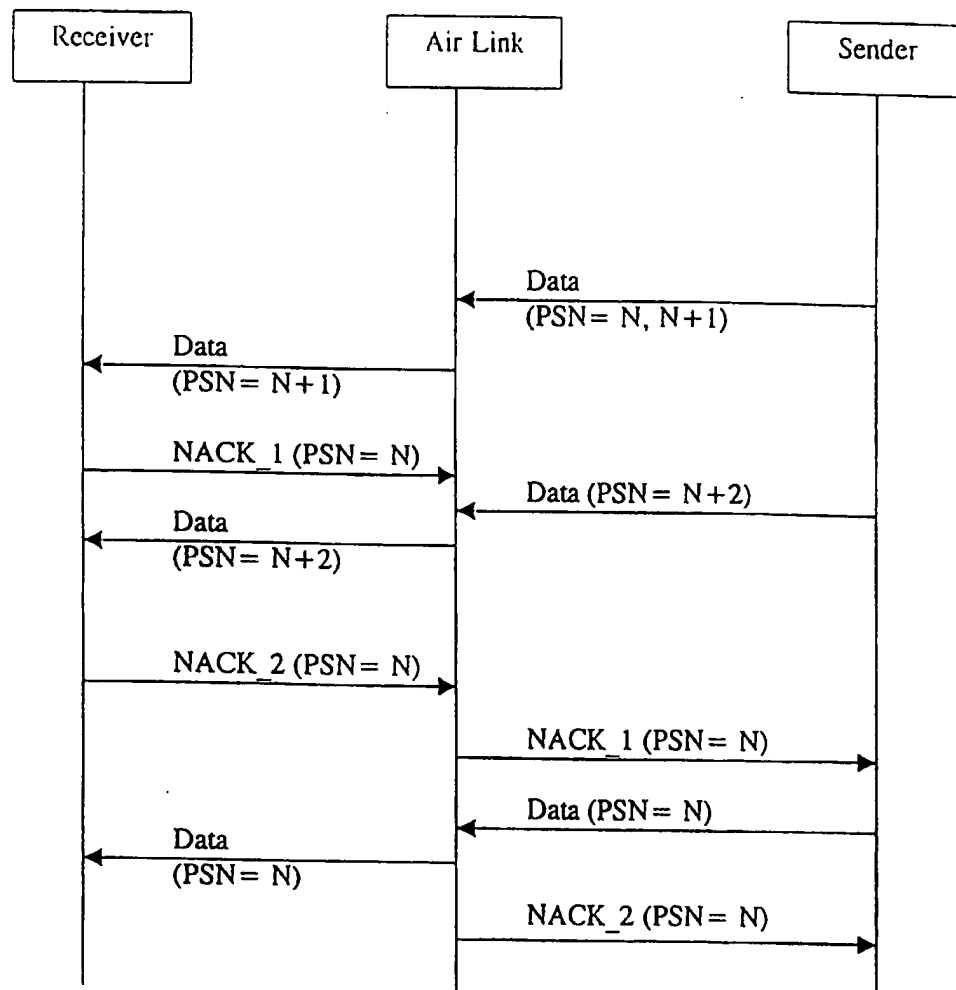


Fig. 4



INTERNATIONAL SEARCH REPORT

Int. Application No
PCT/IB 00/01746

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04L1/16 H04L1/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 695 053 A (AT & T CORP) 31 January 1996 (1996-01-31) abstract column 1, line 32 -column 2, line 17 column 3, line 31 -column 4, line 27 column 5, line 33 -column 6, line 42 column 9, line 31 -column 10, line 8 figures 1,3	1-14
X	US 5 222 061 A (DOSHI BHARAT T ET AL) 22 June 1993 (1993-06-22) abstract column 1, line 62 -column 2, line 14 column 3, line 7 -column 4, line 58 --- -/--	1-6,9, 11-14

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

2 March 2001

Date of mailing of the international search report

09/03/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Toumpoulidis, T

INTERNATIONAL SEARCH REPORT

Int. .lonal Application No

PCT/IB 00/01746

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>GB 2 318 036 A (HITACHI LTD)</p> <p>8 April 1998 (1998-04-08)</p> <p>page 2, line 27 -page 4, line 2</p> <p>figure 2</p> <p>-----</p>	<p>1-6, 11,</p> <p>12</p>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 00/01746

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0695053 A	31-01-1996	US 5570367 A CA 2145782 A,C CN 1123980 A JP 8065355 A	29-10-1996 29-01-1996 05-06-1996 08-03-1996
US 5222061 A	22-06-1993	NONE	
GB 2318036 A	08-04-1998	CA 2216980 A JP 10164176 A US 6034962 A	04-04-1998 19-06-1998 07-03-2000